USING LANDSAT TIME-SERIES AND LIDAR TO INFORM ABOVEGROUND CARBON BASELINE ESTIMATION IN MINNESOTA

Ram K. Deo[1], Grant M. Domke[2], Matthew B. Russell[3], Christopher W. Woodall[2], and Michael J. Falkowski[4]

Abstract—Landsat data has long been used to support forest monitoring and management decisions despite the limited success of passive optical remote sensing for accurate estimation of structural attributes such as aboveground biomass. The archive of publicly available Landsat images dating back to the 1970s can be used to predict historic forest biomass dynamics. In addition, increasing regional scale availability and high sensitivity of LiDAR for biomass mapping also needs exploration of its utility in back-projection modeling. This study has combined recent national forest inventory (NFI) data (2007-2011) with the Landsat data from 1986-2011 and a regional LiDAR dataset acquired by the Minnesota Department of Natural Resources (DNR) to assess the potential of the remote sensing data in predicting aboveground forest biomass back to the 1990 baseline used in the United Nations Framework Convention on Climate Change reporting in the US. Since obtaining cloud-free Landsat images at required seasons for a regional or national study is unlikely, pixel level polynomial models were fitted to a suite of time-series predictors obtained from cloud-free Landsat data of a single scene in Minnesota such that each predictor represented only one growing season between 1986 and 2011. Similarly, selected LiDAR variables were back-projected using Landsat metrics as explanatory variables. The rational for this effort was to obtain a wall-to-wall inventory for any target year that does not have remote sensing data by combining a set of projected predictors and current NFI data. Several candidate models were developed to produce biomass maps for the year 2000 to compare the outputs with the extant map of National Biomass and Carbon Dataset (NBCD) circa 2000 and annual NFI plot measurements. We found that the model including back-projected LiDAR metrics did not significantly improve the prediction accuracy as compared to the model based only on projected Landsat metrics. As the polynomial-projected Landsat-based model provided accuracy similar to the NBCD model, the former may be used for reference mapping back to 1990.

INTRODUCTION

Regional scale, spatially explicit and periodic quantification of aboveground biomass (AGB) is critical for forest carbon accounting and analysis of growth dynamics (Powell et al., 2010). Additionally, a back-in-time biomass baseline is necessary to evaluate national efforts (e.g., forestry-based) on greenhouse gas (GHG) emissions reduction implemented within the United Nations Framework Convention on Climate Change (UNFCCC). Any spatial inventory of forest AGB for the past that lacked sufficient field samples can most reliably apply historic satellite imagery (Huang et al., 2010). Landsat remotely sensed data has long been used to support forest inventory despite limited success of the passive optical data for accurate estimation of AGB. The archive of publicly available Landsat data dating back to the 1970s can be integrated with available standard national

[1] Postdoctoral Research Associate, Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; rkdeo@umn.edu

[2] Research Foresters, Northern Research Station, Forest Inventory and Analysis, USDA Forest Service, St. Paul, MN 55108; gmdomke@fs.fed.us; cwoodall@fs.fed.us

[3] Assistant Professor, Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; russellm@umn.edu

[4] Research Associate Professor, Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; mfalkows@umn.edu

forest inventory (NFI) data to predict biomass in a temporally consistent approach. The NFI system in the US, implemented by the Forest Inventory and Analysis (FIA) program of the USDA Forest Service, has evolved over time with a nationally consistent design adopted since1999. It has been documented that resource estimates based on the current annual design compared to previous periodic designs produce inconsistent results (Goeking et al., 2015).

LiDAR technology has been found to provide the most sensitive remote sensing metrics (e.g., height distribution, strata density, canopy cover) to characterize forest structural attributes. Several studies have highlighted the strengths of LiDAR for landscape scale forest inventory and mapping, and such applications are receiving increasing attention, especially when regional scale LiDAR acquisitions are publicly available (e.g., MN High Resolution Elevation Mapping Project; see http://arcgis.dnr.state.mn.us/maps/mntopo/). This provides an opportunity to combine one-time LiDAR data with the time-series Landsat data for back-projection modeling of AGB.

This study was initially designed to combine a recent cycle (2007-2011) of FIA data with time-series Landsat data from 1986-2011 to assess the efficacy of the optical remote sensing data in back-projecting AGB to the 1990 baseline used in the US for national GHG inventory (NGHGI) reporting to the UNFCCC. An additional goal was to evaluate the inclusion of back-projected LiDAR metrics (as predictors in the modeling frames) from the recently acquired dataset with anticipation of improving the prediction accuracy of AGB for the reference year.

METHODS

FIA Data

Aboveground biomass data for the annual NFI plots measured in 2000 and 2007 to 2011 in northeastern Minnesota were obtained from the FIA program. The data were processed at the FIA, Northern Research Station to comply with the privacy requirements of actual plot locations. The plot biomass data scaled to tons per ha were based on nationally consistent allometric equations (Jenkins et al. 2003) applied to the records of all subplots and micro-plots in each NFI plot.

Remote Sensing Data

We acquired a time-series (1986-2011) of near-anniversary date Landsat-5 Thematic Mapper (TM) surface reflectance data for a single scene in Minnesota (WRS-2 path 27, row 27) from the USGS Climate Data Record (CDR, http://espa.cr.usgs.gov/ordering/new). The acquired images were radiometrically and atmospherically preprocessed at the source via LEDAPS software (http://landsat.usgs.gov/CDR_LSR.php). The time-series collection contained one cloud-free image per peak leaf-on season between mid-July and mid-September when consistent landscape conditions and phenology can be expected due to similar solar geometry; however, only 17 of the 26 seasons contained cloud-free images with a maximum gap of 2 years. The CDR products included surface reflectance-derived spectral indices (http://landsat.usgs.gov/-CDR_ECV.php) as well as individual bands for each acquisition. Six spatial predictors from Landsat data were considered for AGB modeling: Band-5, NDVI (normalized difference vegetation index), NBR (normalized burn ratio), IFZ (integrated forest z-score), TCA (tasseled cap angle), and DI (disturbance index). Band-5, NDVI, and NBR were obtained directly from CDR while IFZ, TCA, and DI were derived as described in Huang et al. (2010), Pflugmacher et al. (2012) and Healey et al. (2005) respectively.

A highly accurate LiDAR dataset (5 cm vertical error), acquired in spring 2011 (May, 3-26) with 1-1.5 m pulse spacing, is publicly available for over 75 percent coverage of the target Landsat scene to the northeastern side called the Arrowhead region. The raw LiDAR data were downloaded from the MnGeo web-portal (http://www.mngeo.state.mn.us/chouse/elevation/-lidar.html#data) and processed to obtain 30 grid-metrics representing canopy cover, elevation distributions, and proportion of returns in vertical strata following Falkowski et al. (2010). The analysis for spatial inventory was focused only to the Arrowhead region of Minnesota.

Modeling Approach

Since obtaining cloud-free Landsat images at nominal intervals for a regional or national study was unlikely, a pixel-level polynomial (3rd degree) curve fit (De Jager and Fox, 2013) was applied to each of the six time-series predictors obtained from the Landsat time-series (17 images between 1986-2011) for the target scene (WRS-2 path-27, row-27). The rationale for this approach was to obtain a wall-to-wall inventory for any target year that does not have cloud-free satellite images by combining a set of projected predictors from polynomial models and current FIA data. The FIA plot data was attached to the Landsat and LiDAR predictors to obtain a reference frame for modelling. The collinear spatial variables were pruned and then best-subset and Random Forest (RF)-based variable selection approach was followed to develop robust and parsimonious spatial models for predicting AGB (Falkowski et al., 2009, 2010). Several AGB models were formulated using the RF-based *k*-Nearest Neighbor (*k*-NN) imputation approach (Crookston and Finley, 2008). The candidate models were dependent on different combinations of Landsat and LiDAR derived spatial predictors, number of observations (plots within years) used in the reference frame and number of

Table 1—Fitted models for aboveground biomass and accuracy statistics for the Arrowhead region in northeastern Minnesota

Model	Predictors and reference years for the model frame	No. of plots	Value of k	% variance explained	Plot-level validation with FIA data in 2000 (n= 262)		Polygon-level validation with NBCD 2000 (n= 110)	
					Bias %	RMSE (mt/ha)	Bias %	RMSE (mt)
1	6 actual TM metrics[Ψ] from 2007, 08, 10 & 11	1347	1	19.27	-2.2432	61.9552	-13.4873	1421.7299
2	6 projected TM metrics from 2007-2011	1661	1	25.79	-1.3484	63.6090	-9.7693	1236.7107
3	TM band-5 and 3 LiDAR[£] metrics from 2011 only	253	1	62.82	-2.1143	61.8594	-13.0971	1417.8929
4	6 projected TM metrics from 2011 only	327	1	24.71	3.5179	58.5749	5.1847	1237.4841
5	6 actual TM metrics from 2007, 08, 10 & 11	1347	3	18.95	-3.1665	61.4549	-13.4693	1427.2498
6	6 projected TM metrics from 2007-2011	1661	3	26.03	-2.7252	63.5456	-10.2685	1263.1804
7	TM band-5 and 3 LiDAR metrics from 2011 only	253	3	62.86	-1.2213	61.5482	-12.9604	1412.8583
8	6 projected TM metrics from 2011 only	327	3	24.83	4.3745	58.0422	4.9054	1237.3781
9	6 actual TM metrics from 2007, 08, 10 & 11	1347	5	19.16	-2.1723	62.3220	-12.9010	1397.7346
10	6 projected TM metrics from 2007-2011	1661	5	25.87	-3.2511	63.8337	-10.0981	1252.0701
11	TM band-5 and 3 LiDAR metrics from 2011 only	253	5	62.8	-1.8642	61.6539	-13.0042	1415.1248
12	6 projected TM metrics from 2011 only	327	5	24.88	4.8440	58.6628	5.5991	1249.3276
NBCD Model					5.0480	43.0694		

[Ψ] 6 TM metrics: Band-5, DI, NBR, IFZ, TCA, and NDVI.
£ 3 LiDAR metrics: Maximum elevn, average elevn, and canopy cover based on percentage of all returns above 2 m.

nearest neighbors (*k*) considered for the imputation (Table 1). For LiDAR dependent models obtained from the plot data and coinciding LiDAR metrics of the acquisition year, only the selected LiDAR metrics were back-projected for a target year via Landsat variables. The selected LiDAR metrics were projected using the RF-based *k*-NN imputation models fitted to a frame obtained from 5000 arbitrary points across the target area where both LiDAR metrics as response and Landsat metrics as predictors were extracted in a GIS environment. The accuracy of model predictions for the year 2000 was evaluated at plot-level using the FIA data of 2000 and also at 110 arbitrary polygon-level (~10 - 133 ha) using an extant AGB map circa 2000 from the National Biomass and Carbon Dataset (NBCD, http://whrc.org/mapping/nbcd/). The performance of AGB models was assessed using statistical measures of bias (predicted - observed) and root mean square error (RMSE), to select the most suitable model for spatial inventory in 1990.

RESULTS AND DISCUSSION

Polynomial curve-fitting to the time-series actual Landsat-derived metrics revealed a better coefficient of determination (R^2) (i.e., temporal consistency) with band-5 where almost 50% of pixels in the target area attained an $R^2 > 0.40$; DI, IFZ, NBR, NDVI and TCA had 37.17%, 32.86%, 31.43%, 17.20% and 4.20% of their respective pixels with $R^2 > 0.40$. The RF-based variable selection algorithm for Landsat dependent models did not identify any collinear metrics but identified only three prime metrics for the LiDAR dependent model. When spatial models of the selected LiDAR metrics dependent on Landsat metrics were developed, reasonable R^2 values were obtained (0.56, 0.49 and 0.65 for ElevMax, ElevAv and Cover-above-2 m, respectively) with the fitting dataset for the year 2011. However, performance of these models when applied for back-projection using the Landsat metrics for the year 2000, were not tested in absence of data.

The plot level validation of AGB prediction using FIA measurements from the inventory year 2000 showed that the model including LiDAR metrics and the projected TM band-5, yield least bias with k = 3 NN in the imputation. Further, the bias of the model including LiDAR was very close to the model dependent on polynomial-projected TM metrics. However, the inspection of RMSE infers that the model based on projected TM predictors from the year 2011 with only 327 plots provided the least error. Additionally, the models based on LiDAR metrics have similar RMSE as the models based only on actual TM derived predictors. Although the LiDAR model performed well when applied in the same year from which it was built, the back-projection was impaired because ultimately it relied on TM predictors which become insensitive in high biomass areas. All the models provided negative bias, except the projected TM only model with fewer plots, suggesting that the imputation models result in under predictions of AGB. This fact of under prediction and the range of observed RMSE are also highlighted in Powell et al. (2013). An interesting finding is that the NBCD model provided the highest bias (but least RMSE) at plot-level compared to all the models formulated in this study. A comparison of polygon-level total estimates by the models evaluated in this study at k= 3 against the NBCD are shown in Figure 1.

CONCLUSIONS

Including current LiDAR data for back-projection of AGB did not improve prediction accuracy. The model based only on back-projected TM, or based on back-projected LiDAR provided similar estimates and hence either could be used. That said, it may be more efficient to just apply projected Landsat metrics rather than exploring many LiDAR metrics and conducting their back projection using Landsat variables. Rather than applying back-projected LiDAR explained by TM variables, it may be better to directly use back-projected TM variables in the model to minimize bias.

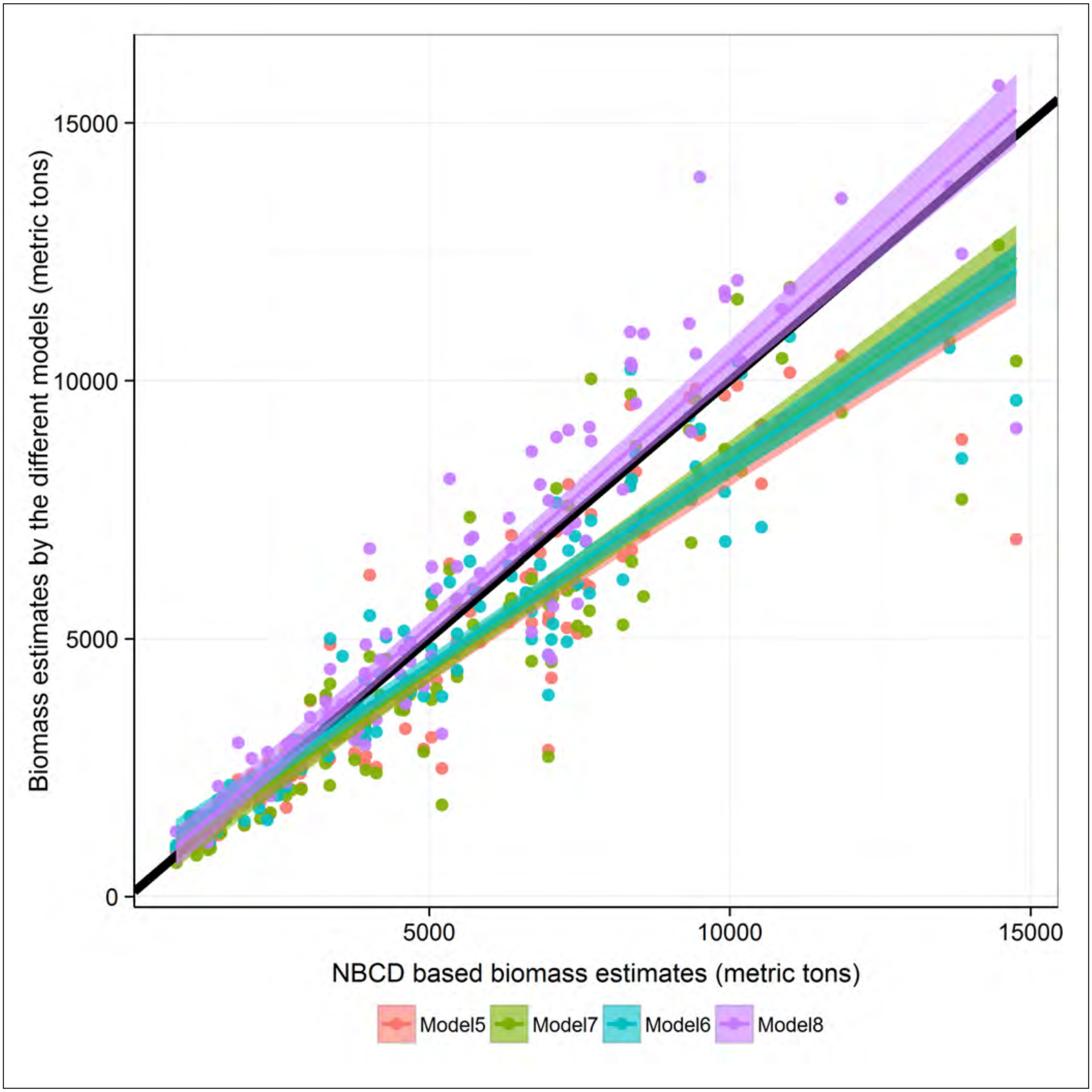

Figure 1—Polygon-level total aboveground biomass estimates for different models (see Table 1) compared with NBCD model estimates for the Arrowhead region in northeastern Minnesota.

LITERATURE CITED

Crookston, N.L.; Finley, A.O. 2008. yaImpute: an R package for kNN imputation. Journal of Statistical Software 23(10): 1-16.

De Jager, N. R.; Fox, T. J. 2013. Curve Fit: a pixel-level raster regression tool for mapping spatial patterns. Methods in Ecology and Evolution 4, 789-792.

Falkowski, M.J.; Evans, J.S.; Martinuzzi, S.; Gessler, P.E.; Hudak, A.T. 2009. Characterizing forest succession with LiDAR data: an evaluation for the inland northwest, U.S.A. Remote Sensing of Environment 113(5): 946-956.

Falkowski, M.J.; Hudak, A.T.; Crookston, N.L.; Gessler, P.E.; Uebler, E.H.; Smith, M.S. 2010. Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating LiDAR data. Canadian Journal of Forest Research 40(2): 184-199.

Goeking, S. 2015. Disentangling forest change from forest inventory change: A case study from the US interior west. Journal of Forestry 113 (*): In press. http://dx.doi.org/10.5849/jof.14-088.

Healey, S.P.; Yang, Z.; Cohen, W. B.; Pierce, D. J. 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. Remote Sensing of Environment 97 (3), 301-310.

Huang, C.; Goward, S. N.; Masek, J. G.; Thomas, N.; Zhu, Z.; Vogelmann, J. E. 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sensing of Environment 114 (1), 183-198.

Jenkins, J. C.; Chojnacky, D. C.; Heath, L. S.; Birdsey, R. A. 2003. National-scale biomass estimators for United States tree species. Forest Science 49 (1): 12-35.

Pflugmacher, D; Cohen, W. B; Kennedy, R. E. 2012. Using Landsat-derived disturbance history (1972-2010) to predict current forest structure. Remote Sensing of Environment 122 (Landsat Legacy Special Issue), 146-165.

Powell, S.L.; Cohen, W.B.; Healey, S.P.; Kennedy, R.E.; Moisen, G.G.; Pierce, K.B.; Ohmann, J.L. 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modelling approaches. Remote Sensing of Environment 114(5): 1053-1068.

Powell, S. L.; Cohen, W. B.; Kennedy, R. E.; Healey, S. P.; Huang, C. 2013. Observation of trends in biomass loss as a result of disturbance in the conterminous U.S.: 1986-2004. Ecosystems. Springer. ISSN 1432-9840. DOI 10.1007/s10021-013-9713-9.